CHAPTER 6

- JOINTLY DISTRIBUTED RANDOM VARIABLES

6.1 - Joint distribution functions

Thus far, we have only seen probability distributions for single random variables.

→ However we are often interested in probability statements concerning 2 or more random variables.

In order to deal with such probabilities we define, for any 2 random variables X and Y, the joint cumulative probability distribution function of X & Y by:

$$\boxed{* \; F(a,b) = P(X \le a, Y \le b) \quad -\infty < a,b < \infty}$$

The distribution of X can be obtained from the joint distribution of X & Y as follows

$$F_X(a) = P(X \le a)$$
$$= P(X \le a, Y < \infty)$$
$$= P\left(\lim_{b \to \infty} (X \le a, Y \le b)\right)$$
$$= \lim_{b \to \infty} P(X \le a, Y \le b)$$
$$= \lim_{b \to \infty} F(a,b)$$
$$= F(a, \infty)$$

Note:

The preceding set of equalities, we have once made use of the fact that probability is a continuous set. [That is, event] function.

- Cumulative distribution function of Y.

$$F_Y(b) = P(Y \le b)$$

$$= \lim_{a \to \infty} F(a, b)$$

$$= F(\infty, b)$$

- The distribution functions F_X & F_Y are sometimes reffered to as marginal distributions of X & Y.

- All joint probability statements about X & Y can, in theory, be answered in terms of their joint distribution function.

- **For instance:**

Suppose we wanted to compute the joint probability that X is greater than a & Y is greater than b.

i.e.

$$P(X > a, Y > b) = 1 - P[(X > a, Y > b)^c]$$

$$= 1 - P((X > a)^c \cup P(Y > b)^c) \quad (1.1)$$

$$= 1 - P((X \leq a) \cup P(Y \leq b))$$

$$= 1 - [P(X \leq a) + P(Y \leq b) - P(X \leq a, Y \leq b)]$$

$= 1 - F_X(a) - F_Y(b) + F(a,b)$

$\therefore$

$P(a_1 < X < a_2, b_1 < Y \leq b_2)$

$= F(a_2,b_2) + F(a_1,b_1) - F(a_1,b_2) - F(a_2,b_1)$ (1.2)

where $a_1 < a_2, b_1 < 2$

In the case when X & Y are both discrete random variables, it is convenient to define the joint probability mass function of X & Y.

$$p(x,y) = P(X=x, Y=y)$$

The probability mass function of X can be obtained from $p(x,y)$ by

$$p_X(x) = P\{X = x\} = \sum_{y:p(x,y)>0} p(x,y)$$

Similarly,

$$p_Y(y) = \sum_{x:p(x,y)>0} p(x,y)$$

EXAMPLE 1c

The joint density function of X and Y is given by

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty,\ 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute (a) $P\{X > 1, Y < 1\}$, (b) $P\{X < Y\}$, and (c) $P\{X < a\}$.

Solution. (a)

$$\begin{aligned} P\{X > 1, Y < 1\} &= \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}\,dx\,dy \\ &= \int_0^1 2e^{-2y}\left(-e^{-x}\Big|_1^\infty\right) dy \\ &= e^{-1}\int_0^1 2e^{-2y}\,dy \\ &= e^{-1}(1 - e^{-2}) \end{aligned}$$

(b)

$$\begin{aligned} P\{X < Y\} &= \iint_{(x,y):x<y} 2e^{-x}e^{-2y}\,dx\,dy \\ &= \int_0^\infty \int_0^y 2e^{-x}e^{-2y}\,dx\,dy \end{aligned}$$

$$\begin{aligned} &= \int_0^\infty 2e^{-2y}(1 - e^{-y})dy \\ &= \int_0^\infty 2e^{-2y}dy - \int_0^\infty 2e^{-3y}dy \\ &= 1 - \frac{2}{3} \\ &= \frac{1}{3} \end{aligned}$$

(c)

$$\begin{aligned} P\{X < a\} &= \int_0^a \int_0^\infty 2e^{-2y}e^{-x}\,dy\,dx \\ &= \int_0^a e^{-x}\,dx \\ &= 1 - e^{-a} \end{aligned}$$

■

* EXAMPLE 1f - The multinomial distribution

- One of the most important joint distributions which arises when a sequence of n-independent and identical experiments is performed

- Suppose that each experiment can result in any one of r possible outcomes, with respective probabilities $p_1, p_2, p_3, \ldots, p_r$, $\sum_{i=1}^{r} p_i = 1$.

If we let X_i denote the number of the n-experiments that result in outcome number i. then

$$P\{X_1 = n_1, X_2 = n_2, \ldots, X_r = n_r\} = \frac{n!}{n_1!n_2!\cdots n_r!}p_1^{n_1}p_2^{n_2}\cdots p_r^{n_r} \tag{1.5}$$

whenever $\sum_{i=1}^{r} n_i = n$.

↳ The joint distribution whose probability mass function is specified by Equation (1.5) is called multinomial distribution.

- When $r=2$, the multinomial distribution reduces to binomial distribution

As an application of the multinomial distribution, suppose that a fair die is rolled 9 times. The probability that 1 appears three times, 2 and 3 twice each, 4 and 5 once each, and 6 not at all is

(n_i)

$$\frac{9!}{3!2!2!1!1!0!}\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^0 = \frac{9!}{3!2!2!}\left(\frac{1}{6}\right)^9$$

■

6.2 - INDEPENTENT RANDOM VARIABLES

The random variables X and Y are said to be <u>independent</u> if, for any 2 sets of real numbers A & B

$$P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$$

↳ In other words, X & Y are independent if for all A & B, the events $E_A = (X \in A)$ & $F_B = (Y \in B)$ are independent.

* Proof can be found on pg 256.

EXAMPLE 2c (Interesting Example)

A man and a woman decide to meet at a certain location. If each of them independently arrives at a time uniformly distributed between 12 noon and 1 P.M., find the probability that the first to arrive has to wait longer than 10 minutes.

Solution. If we let X and Y denote, respectively, the time past 12 that the man and the woman arrive, then X and Y are independent random variables, each of which is uniformly distributed over $(0, 60)$. The desired probability, $P\{X + 10 < Y\} + P\{Y + 10 < X\}$, which, by symmetry, equals $2P\{X + 10 < Y\}$, is obtained as follows:

$$
\begin{aligned}
2P\{X + 10 < Y\} &= 2 \iint_{x+10<y} f(x, y)\, dx\, dy \\
&= 2 \iint_{x+10<y} f_X(x) f_Y(y)\, dx\, dy \\
&= 2 \int_{10}^{60} \int_{0}^{y-10} \left(\frac{1}{60}\right)^2 dx\, dy \\
&= \frac{2}{(60)^2} \int_{10}^{60} (y - 10)\, dy \\
&= \frac{25}{36}
\end{aligned}
$$

■

Our next example presents the oldest problem dealing with geometrical probabilities. It was first considered and solved by Buffon, a French naturalist of the 18th century, and is usually referred to as *Buffon's needle problem*.

~~#~~ SOS!

EXAMPLE 2d ***Buffon's needle problem***

A table is ruled with equidistant parallel lines a distance D apart. A needle of length L, where $L \leq D$, is randomly thrown on the table. What is the probability that the needle will intersect one of the lines (the other possibility being that the needle will be completely contained in the strip between two lines)?

Solution. Let us determine the position of the needle by specifying (1) the distance X from the middle point of the needle to the nearest parallel line and (2) the angle θ between the needle and the projected line of length X. (See Figure 6.2.) The needle will intersect a line if the hypotenuse of the right triangle in Figure 6.2 is less than $L/2$—that is, if

$$\frac{X}{\cos\theta} < \frac{L}{2} \quad \text{or} \quad X < \frac{L}{2}\cos\theta$$

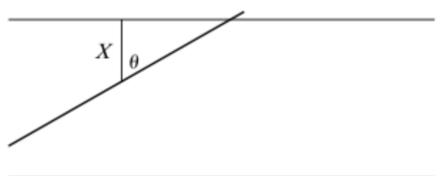

FIGURE 6.2

As X varies between 0 and $D/2$ and θ between 0 and $\pi/2$, it is reasonable to assume that they are independent, uniformly distributed random variables over these respective ranges. Hence,

$$\begin{aligned} P\left\{X < \frac{L}{2}\cos\theta\right\} &= \iint\limits_{x<L/2\cos y} f_X(x) f_\theta(y)\,dx\,dy \\ &= \frac{4}{\pi D}\int_0^{\pi/2}\int_0^{L/2\cos y} dx\,dy \\ &= \frac{4}{\pi D}\int_0^{\pi/2}\frac{L}{2}\cos y\,dy \\ &= \frac{2L}{\pi D} \end{aligned}$$

■

More examples can be found on

pg 259.

- How can a computer choose a random subset.

Most computers are able to generate the value of, or *simulate*, a uniform (0, 1) random variable by means of a built-in subroutine that (to a high degree of approximation)

produces such "random numbers." As a result, it is quite easy for a computer to simulate an indicator (that is, a Bernoulli) random variable. Suppose I is an indicator variable such that

$$P\{I = 1\} = p = 1 - P\{I = 0\}$$

The computer can simulate I by choosing a uniform (0, 1) random number U and then letting

$$I = \begin{matrix} 1 & \text{if } U < p \\ 0 & \text{if } U \geq p \end{matrix}$$

Suppose that we are interested in having the computer select $k, k \leq n$, of the numbers $1, 2, \ldots, n$ in such a way that each of the $\binom{n}{k}$ subsets of size k is equally likely to be chosen. We now present a method that will enable the computer to solve this task. To generate such a subset, we will first simulate, in sequence, n indicator variables $I_1, I_2, \ldots, I_n$, of which exactly k will equal 1. Those i for which $I_i = 1$ will then constitute the desired subset.

To generate the random variables $I_1, \ldots, I_n$, start by simulating n independent uniform (0, 1) random variables $U_1, U_2, \ldots, U_n$. Now define

$$I_1 = \begin{cases} 1 & \text{if } U_1 < \dfrac{k}{n} \\ 0 & \text{otherwise} \end{cases}$$

and then, once $I_1, \ldots, I_i$ are determined, recursively set

$$I_{i+1} = \begin{cases} 1 & \text{if } U_{i+1} < \dfrac{k - (I_1 + \cdots + I_i)}{n - i} \\ 0 & \text{otherwise} \end{cases}$$

In words, at the $(i+1)$th stage we set I_{i+1} equal to 1 (and thus put $i+1$ into the desired subset) with a probability equal to the remaining number of places in the subset $\left(\text{namely}, k - \sum_{j=1}^{i} I_j\right)$, divided by the remaining number of possibilities (namely, $n - i$). Hence, the joint distribution of $I_1, I_2, \ldots, I_n$ is determined from

$$P\{I_1 = 1\} = \frac{k}{n}$$

$$P\{I_{i+1} = 1 | I_1, \ldots, I_i\} = \frac{k - \sum_{j=1}^{i} I_j}{n - i} \qquad 1 < i < n$$

The proof that the preceding formula results in all subsets of size k being equally likely to be chosen is by induction on $k + n$. It is immediate when $k + n = 2$ (that is, when $k = 1, n = 1$), so assume it to be true whenever $k + n \le l$. Now, suppose that $k + n = l + 1$, and consider any subset of size k—say, $i_1 \le i_2 \le \cdots \le i_k$—and consider the following two cases.

Case 1: $i_1 = 1$

$$\begin{aligned} &P\{I_1 = I_{i_2} = \cdots = I_{i_k} = 1, I_j = 0 \text{ otherwise}\} \\ &\quad = P\{I_1 = 1\} P\{I_{i_2} = \cdots = I_{i_k} = 1, I_j = 0 \text{ otherwise} | I_1 = 1\} \end{aligned}$$

Now given that $I_1 = 1$, the remaining elements of the subset are chosen as if a subset of size $k - 1$ were to be chosen from the $n - 1$ elements $2, 3, \ldots, n$. Hence, by the induction hypothesis, the conditional probability that this will result in a given subset of size $k - 1$ being selected is $1/\binom{n-1}{k-1}$. Hence,

$$\begin{aligned} &P\{I_1 = I_{i_2} = \cdots = I_{i_k} = 1, I_j = 0 \text{ otherwise}\} \\ &\quad = \frac{k}{n} \frac{1}{\binom{n-1}{k-1}} = \frac{1}{\binom{n}{k}} \end{aligned}$$

Case 2: $i_1 \ne 1$

$$\begin{aligned} &P\{I_{i_1} = I_{i_2} = \cdots = I_{i_k} = 1, I_j = 0 \text{ otherwise}\} \\ &\quad = P\{I_{i_1} = \cdots = I_{i_k} = 1, I_j = 0 \text{ otherwise} | I_1 = 0\} P\{I_1 = 0\} \\ &\quad = \frac{1}{\binom{n-1}{k}} \left(1 - \frac{k}{n}\right) = \frac{1}{\binom{n}{k}} \end{aligned}$$

where the induction hypothesis was used to evaluate the preceding conditional probability.

Thus, in all cases, the probability that a given subset of size k will be the subset chosen is $1/\binom{n}{k}$. ■

Remark. The foregoing method for generating a random subset has a very low memory requirement. A faster algorithm that requires somewhat more memory is presented in Section 10.1. (The latter algorithm uses the last k elements of a random permutation of $1, 2, \ldots, n$.) ■

EXAMPLE 2i

- Probabilistic interpretation of half-life.

- Let $N(t)$ denote the number of nuclei contained in a radioactive mass of material at time t.

- For some value h (half-life)

$$N(t) = 2^{-t/h} N(0) \quad t \geq 0$$

- For any non-negative s and t.

$$N(t+s) = 2^{-(s+t)/h} N(0) = 2^{-t/h} N(s)$$

it follows that no matter how much time s has already elapsed, in addition time t, the number of existing nuclei will decrease by the factor $2^{-t/h}$

Probabilistic interpretation of the half-life h: The lifetimes of the individual nuclei are independent random variables having a life distribution that is exponential with median equal to h. That is, if L represents the lifetime of a given nucleus, then

$$P\{L < t\} = 1 - 2^{-t/h}$$

(Because $P\{L < h\} = \frac{1}{2}$ and the preceding can be written as

$$P\{L < t\} = 1 - \exp\left\{-t\frac{\log 2}{h}\right\}$$

it can be seen that L indeed has an exponential distribution with median h.)

Note that, under the probabilistic interpretation of half-life just given, if one starts with $N(0)$ nuclei at time 0, then $N(t)$, the number of nuclei that remain at time t, will have a binomial distribution with parameters $n = N(0)$ and $p = 2^{-t/h}$. Results of Chapter 8 will show that this interpretation of half-life is consistent with the deterministic model when considering the proportion of a large number of nuclei that decay over a given time frame. However, the difference between the deterministic and probabilistic interpretation becomes apparent when one considers the actual number of

decayed nuclei. We will now indicate this with regard to the question of whether protons decay.

There is some controversy over whether or not protons decay. Indeed, one theory predicts that protons should decay with a half-life of about $h = 10^{30}$ years. To check this prediction empirically, it has been suggested that one follow a large number of protons for, say, one or two years and determine whether any of them decay within that period. (Clearly, it would not be feasible to follow a mass of protons for 10^{30} years to see whether one-half of it decays.) Let us suppose that we are able to keep track of $N(0) = 10^{30}$ protons for c years. The number of decays predicted by the deterministic model would then be given by

$$\begin{aligned}
N(0) - N(c) &= h(1 - 2^{-c/h}) \\
&= \frac{1 - 2^{-c/h}}{1/h} \\
&\approx \lim_{x \to 0} \frac{1 - 2^{-cx}}{x} \qquad \text{since } \frac{1}{h} = 10^{-30} \approx 0 \\
&= \lim_{x \to 0} (c2^{-cx} \log 2) \qquad \text{by L'Hôpital's rule} \\
&= c \log 2 \approx .6931c
\end{aligned}$$

For instance, the deterministic model predicts that in 2 years there should be 1.3863 decays, and it would thus appear to be a serious blow to the hypothesis that protons decay with a half-life of 10^{30} years if no decays are observed over those 2 years.

Let us now contrast the conclusions just drawn with those obtained from the probabilistic model. Again, let us consider the hypothesis that the half-life of protons is $h = 10^{30}$ years, and suppose that we follow h protons for c years. Since there is a huge number of independent protons, each of which will have a very small probability of decaying within this time period, it follows that the number of protons which decay will have (to a very strong approximation) a Poisson distribution with parameter equal to $h(1 - 2^{-c/h}) \approx c \log 2$. Thus,

$$\begin{aligned}
P\{0 \text{ decays}\} &= e^{-c \log 2} \\
&= e^{-\log(2^c)} = \frac{1}{2^c}
\end{aligned}$$

and, in general,

$$P\{n \text{ decays}\} = \frac{2^{-c}[c \log 2]^n}{n!} \qquad n \geq 0$$

Thus we see that even though the average number of decays over 2 years is (as predicted by the deterministic model) 1.3863, there is 1 chance in 4 that there will not be any decays, thereby indicating that such a result in no way invalidates the original hypothesis of proton decay. ■

Remark. *Independence is a symmetric relation.* The random variables X and Y are independent if their joint density function (or mass function in the discrete case) is the product of their individual density (or mass) functions. Therefore, to say that X is independent of Y is equivalent to saying that Y is independent of X—or just that X and Y are independent. As a result, in considering whether X is independent of Y in situations where it is not at all intuitive that knowing the value of Y will not change the probabilities concerning X, it can be beneficial to interchange the roles of

6.3 - SUMS OF INDEPENTENT RANDOM VARIABLES

It is often important to be able to calculate the distribution of $X+Y$ from the distributions of X and Y when X & Y are independent.

↳ Suppose X & Y are independent continuous ; random variables having pdf f_X & f_Y

$\therefore$

$$F_{X+Y}(a) = P(X+Y \leq a)$$
$$= \iint_{x+y \leq a} f_X(x) f_Y(y) \, dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)\, f_Y(y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)\,dx\, f_Y(y)\,dy$$

$$= \int_{-\infty}^{\infty} F_X(a-y)\, f_Y(y)\,dy \qquad (3.1)$$

$\therefore$ The cumulative distribution function F_{X+Y} is called convolution of the distributions F_X & F_Y.

By differentiating,

$$f_{X+Y}(a) = \frac{d}{da}\int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\,dy$$
$$= \int_{-\infty}^{\infty} \frac{d}{da} F_X(a-y) f_Y(y)\,dy \qquad (3.2)$$
$$= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y)\,dy$$

6.3.1 - 6.3.5 [pg 278]
Different continuous random variables.

6.4 - Conditional Distributions: Discrete Case

Reminder:
For any 2 events E & F, the conditional probability of E given F is defined by: [provided P(F) > 0]

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Hence if X & Y are discrete random variables, it is natural to define the conditional probability mass function of X given $Y=y$.

$$p_{X|Y}(x|y) = P(X=x \mid Y=y)$$
$$= \frac{P(X=x), P(Y=y)}{P(Y=y)}$$
$$= \frac{p(x,y)}{p_Y(y)}.$$

$\rightarrow$ conditional distribution:

$$F_{X|Y}(x|y) = P(X \le x \mid Y = y)$$

$$= \sum_{a \le x} p_{X|Y}(a|y)$$

In other words, the definitions are exactly the same as in the unconditional case, except that everything is now conditional on the event that $Y = y$. If X is independent of Y, then the conditional mass function and the distribution function are the same as the respective unconditional ones. This follows because if X is independent of Y, then

$$\begin{aligned} p_{X|Y}(x|y) &= P\{X = x|Y = y\} \\ &= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\ &= \frac{P\{X = x\}P\{Y = y\}}{P\{Y = y\}} \\ &= P\{X = x\} \end{aligned}$$

EXAMPLE 4a Simple example

Suppose that $p(x, y)$, the joint probability mass function of X and Y, is given by

$$p(0, 0) = .4 \quad p(0, 1) = .2 \quad p(1, 0) = .1 \quad p(1, 1) = .3$$

Calculate the conditional probability mass function of X given that $Y = 1$.

Solution. We first note that

$$p_Y(1) = \sum_x p(x, 1) = p(0, 1) + p(1, 1) = .5$$

Hence,

$$p_{X|Y}(0|1) = \frac{p(0, 1)}{p_Y(1)} = \frac{2}{5}$$

and

$$p_{X|Y}(1|1) = \frac{p(1, 1)}{p_Y(1)} = \frac{3}{5}$$

■

6.5 - Conditional Distributions: Continuous CASE

If X and Y have a joint probability density function $f(x,y)$, then the conditional probability density function of X given that $Y=y$ is defined, for all values of y such that $f_Y(y) > 0$

By:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

To motivate this definition, multiply the left-hand side by dx and the right-hand side by $(dx\,dy)/dy$ to obtain

$$\begin{aligned} f_{X|Y}(x|y)\,dx &= \frac{f(x,y)\,dx\,dy}{f_Y(y)\,dy} \\ &\approx \frac{P\{x \le X \le x + dx, y \le Y \le y + dy\}}{P\{y \le Y \le y + dy\}} \\ &= P\{x \le X \le x + dx | y \le Y \le y + dy\} \end{aligned}$$

In other words, for small values of dx and dy, $f_{X|Y}(x|y)dx$ represents the conditional probability that X is between x and $x + dx$ given that Y is between y and $y + dy$.

The use of conditional densities allows us to define conditional probabilities of events associated with one random variable when we are given the value of a second random variable.

EXAMPLE 5a

The joint density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{12}{5}x(2-x-y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the conditional density of X given that $Y = y$, where $0 < y < 1$.

Solution. For $0 < x < 1, 0 < y < 1$, we have

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ &= \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\,dx} \\ &= \frac{x(2-x-y)}{\int_0^1 x(2-x-y)\,dx} \\ &= \frac{x(2-x-y)}{\frac{2}{3} - y/2} \\ &= \frac{6x(2-x-y)}{4-3y} \end{aligned}$$

■

The Bivariate Normal Distribution

EXAMPLE 5c The Bivariate Normal Distribution

One of the most important joint distributions is the bivariate normal distribution. We say that the random variables X, Y have a bivariate normal distribution if, for constants $\mu_x, \mu_y, \sigma_x > 0, \sigma_y > 0, -1 < \rho < 1$, their joint density function is given, for all $-\infty < x, y < \infty$, by

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}$$

We now determine the conditional density of X given that $Y = y$. In doing so, we will continually collect all factors that do not depend on x and represent them by the constants C_i. The final constant will then be found by using that $\int_{-\infty}^{\infty} f_{X|Y}(x|y)\,dx = 1$. We have

$$\begin{aligned}
f_{X|Y}(x|y) &= \frac{f(x, y)}{f_Y(y)} \\
&= C_1 f(x, y) \\
&= C_2 \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{x(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\} \\
&= C_3 \exp\left\{-\frac{1}{2\sigma_x^2(1-\rho^2)}\left[x^2 - 2x\left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right)\right]\right\} \\
&= C_4 \exp\left\{-\frac{1}{2\sigma_x^2(1-\rho^2)}\left[x - \left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right)\right]^2\right\}
\end{aligned}$$

Recognizing the preceding equation as a normal density, we can conclude that, given $Y = y$, the random variable X is normally distributed with mean $\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y)$ and variance $\sigma_x^2(1 - \rho^2)$. Also, because the joint density of Y, X is exactly the same as that of X, Y, except that μ_x, σ_x are interchanged with μ_y, σ_y, it similarly follows that the conditional distribution of Y given $X = x$ is the normal distribution with mean $\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x)$ and variance $\sigma_y^2(1 - \rho^2)$. It follows from these results that the necessary and sufficient condition for the bivariate normal random variables X and Y to be independent is that $\rho = 0$ (a result that also follows directly from their joint density, because it is only when $\rho = 0$ that the joint density factors into two terms, one depending only on x and the other only on y).

With $C = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$, the marginal density of X can be obtained from

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x, y)\, dy \\
&= C\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right.\right. \\
&\qquad \left.\left. -2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\} dy
\end{aligned}
$$

Making the change of variables $w = \frac{y-\mu_y}{\sigma_y}$ gives

$$
\begin{aligned}
f_X(x) &= C\sigma_y \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\} \\
&\quad \times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[w^2 - 2\rho\frac{x-\mu_x}{\sigma_x}w\right]\right\} dw \\
&= C\sigma_y \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_x}{\sigma_x}\right)^2(1-\rho^2)\right\} \\
&\quad \times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[w - \rho\frac{x-\mu_x}{\sigma_x}\right]^2\right\} dw
\end{aligned}
$$

Because

$$
\frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[w - \frac{\rho}{\sigma_x}(x-\mu_x)\right]^2\right\} dw = 1
$$

we see that

$$
\begin{aligned}
f_X(x) &= C\sigma_y\sqrt{2\pi(1-\rho^2)}\, e^{-(x-\mu_x)^2/2\sigma_x^2} \\
&= \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-(x-\mu_x)^2/2\sigma_x^2}
\end{aligned}
$$

That is, X is normal with mean μ_x and variance σ_x^2. Similarly, Y is normal with mean μ_y and variance σ_y^2. ■

6.7 - JOINT probability distribution of Functions of Random Variables

(see book)

+ 6.8